PCT/GB00/03576

GB00/3576

REC'D 05 OCT 2000

WIPO PCT

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

10/088417

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

Dated 21 September 2000

nts Form 1/77

Patents Act 1977
(Rule 16)

RECEIVED BY HAND
17 SEP 1999
D
THE PATENT OFFICE

The Patent Office

1/77

20SEP99 E477564-1 D01038
P01/7700 0.00 - 9922013.9

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference JPD/SMH/SUSSEX.1

2. Patent applicat... 9922013.9 17 SEP 1999

3. ...ame, address and postcode of the or of each applicant *(underline all surnames)*

UNIVERSITY OF SUSSEX
Falmer
Brighton
BN1 9QG

Patents ADP number *(if you know it)* 6254866003

If the applicant is a corporate body, give the country/state of its incorporation ENGLAND

4. Title of the invention

PEPTIDES

5. Name of your agent *(if you have one)*

WITHERS & ROGERS

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

Goldings House
2, Hays Lane
London
SE1 2HW

Patents ADP number *(if you know it)* 1776001

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d))

YES

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	17
Claim(s)	3
Abstract	1
Drawing(s)	4 + 4

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Signature [signature] Date 17/09/99

12. Name and daytime telephone number of person to contact in the United Kingdom

J P DEAN - Tel:)11 7 925 3030

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

JPD/Sussex1
O:\SPECS\JPD\SUSSEX1.lwp

Peptides

This invention relates to protein structures and to methods of producing those protein structures and to protein fibres and other materials and assemblies produced using those protein structures.

The process of molecular self-assembly is central to all biological systems and is assuming increasing importance and application in biotechnology (L. Q. Gu, *et al* (1999) *Nature* **398**, 686) and nanotechnology (K. E. Drexler, (1999) TIBTECH **17**, 5). The characterization of natural biomolecular assemblies motivates and directs the development of model self-assembling systems and, in turn, these advance our understanding of biology. For proteins at least, the coiled coil is arguably the simplest self-assembling system. Coiled coils are protein-folding motifs that direct and cement a wide variety of protein-protein interactions (A. Lupas, (1996) *Trends Biochem. Sci* **21**, 375). In structural terms, coiled coils are relatively straightforward: they are α-helical bundles with between 2 and 5 strands that can be arranged in parallel, antiparallel or mixed topologies. The basic sequence features that guide the formation of coiled coils from peptides are reasonably well understood (P. B. Harbury *et al* (1993) *Science* **262**, 1401; D. N. Woolfson and T. Alber (1995) *Protein Sci.* **4**, 1596). For instance, most coiled-coil sequences are dominated by a 7-residue repeat of hydrophobic (H) and polar (P) residues, $(HPPHPPP)_n$, known as the "heptad repeat". When configured into an α-helix this pattern gives an amphipathic structure, the hydrophobic face of which directs oligomer-assembly. Furthermore, both the number and the direction of chains within a coiled-coil bundle is determined predominantly by residues that form or flank the hydrophobic core namely, residues at the first, fourth, fifth and seventh positions of the heptad repeat. For instance, coiled coils which form dimers (i.e. two-stranded assemblies) usually have A isoleucine or valine residue at the first position and a leucine residue at the fourth position. By contrast, coiled coils that form trimers (i.e. three-stranded assemblies) often have the same residues (i.e both isoleucine or both leucine) at both "H" positions. Finally, hetero-oligomers (that is coiled coils made from strands with different amino-acid sequences) may be directed by complementary charged interactions that flank the hydrophobic core. For these reasons, there have been a

number of successful *de novo* protein designs based on the coiled coil These include some ambitious structures that extend the natural repertoire of coiled-coil motifs (S. Nautiyal *et al* (1995) *Biochemistry* **34**, 11645; A. Lombardi *et al* (1996) *Biopolymers* **40**, 495; D. H. Lee *et al* (1996) *Nature* **382**, 525; P. B. Harbury *et al* (1998) *Science* **282**, 1462; J. P. Schneider *et al* (1998) *Folding Des.* **3**, R29).

Other relevant aspects of coiled coil structure are described in WO99/11774, the disclosure of which is incorporated herein by way of reference.

This understanding of coiled coils, and the resulting protein designs, centres on short structures as exemplified by the leucine-zipper motifs (E. K. O'Shea *et al* (1989) *Science* **243**, 538; E. K. O'Shea *et al* (1991) *Science* **254**, 539), which are found in a variety of transcription factors. In contrast, most natural coiled coils extend over hundreds of amino acids (A. Lupas (1996) *Supra*; J. Sodek *et al* (1972) *Proc. Natl. Acad. Sci. U.S.A* **69**, 3800) and many assemble further to form thicker, multi-stranded filaments (H. Herrmann and U. Aebi (1998) *Curr. Opin. Struct. Biol.* **8**, 177).

With the goal of making elongated structures to improve our understanding of coiled coils, and to develop protein-design studies, two 28-residue peptides were initially designed — dubbed Self-Assembling Fibre peptides, SAF-p1 and SAF-p2 — to fold and form extended fibres when mixed. Focusing on the buried, hydrophobic-core positions of the structure, rules were incorporated to direct parallel dimer formation and to guard against alternative oligomers and topologies (P. B. Harbury *et al* (1993) *supra*; D. N. Woolfson and T. Alber (1995) *supra*; L. J. Gonzalez *et al* (1996) *Nature Struct. Biol.* **3**, 1011). The building block of the design was a staggered heterodimer with overhanging or "sticky" ends. This contrasts with and distinguishes it from the natural and designer coiled-coil assemblies that have been characterized to date, in which the polypeptide strands align in-register, i.e they have blunt or "flush" ends. Complementary core interactions and flanking ion-pairs were incorporated into the overhangs to facilitate longitudinal association of the heterodimers (Figs. 1&2). This principle of using "sticky ends" is well developed in molecular biology for assembling DNA (S. J. Palmer *et al* (1998) *Nucleic Acids Res.* **26**, 2560), and has been used to design intricate DNA crystals (E. Winfree *et al* (1998) *Nature* **394**, 539). However, to our knowledge, our application of sticky end-directed molecular assembly to peptides is

new; although we do note that head-to-tail packing of helices has been observed in recently solved crystal structures for two designer peptides (N. L. Ogihara *et al* (1997) *Protein Sci.* **6**, 80; G. G. Prive *et al* (1999) *Protein Sci.* **8**, 1400). These were helical peptides that crystallised with their helical ends in contact so as to form pseudo-continuous helices in the solid state. In other words they formed "blunt-ended" arrangements.

According to one aspect of the invention there is provided a protein structure comprising a plurality of first peptide monomer units arranged in a first strand and a plurality of second peptide monomer units arranged in a second strand, the strands preferably forming a coiled coil structure, and in which a first peptide monomer unit in the first strand extends beyond a corresponding second peptide monomer unit in the second strand in the direction of the strands. The protein structures of the invention have numerous advantages. For example, relatively long protein fibres can be formed with little material - 1 μl of a 100 μM solution of the peptide monomers may provide enough material to form 10 m of continuous fibre 50 nm thick.

At least one charged amino acid residue of the first peptide monomer unit may be arranged to attract an oppositely-charged amino acid residue of the second peptide monomer unit. Preferably, the charged amino acid residue is in an end portion of the first peptide monomer unit, which extends beyond the corresponding second peptide monomer unit in the second strand. At least one strand may consist solely of first or second peptide monomer units respectively i.e homogenous strands. Heterologous strands are also contemplated. The peptide monomer units may comprise a repeating structural unit. Preferably, the repeating structural unit comprises a heptad repeat motif, preferably having the pattern:

abcdefg

the repeat may include isoleucine or asparagine at position a and leucine at position d. Other repeats (e.g hendecads - abcdefghijk) and amino acid compositions may be used (see WO99/11774).

Preferably, the heptad repeat comprises oppositely-charged residues at postions e and g respectively. The oppositely-charged residues may be, for example, glutamic acid and

lysine residues or arginine and aspartic acid. The use of synthetic amino acids, such as ornithine is also envisaged.

A protein structure in accordance with the invention may be also stabilised by pairs of asparagine residues in the "a" positions provided by corresponding first and second peptide monomer units.

In a preferred protein structure, the first and second peptide monomer units have the following sequences:

a) KIAALKQKIASLKQEIDALEYENDALEQ (SAF-p1C) and

b) KIRALKAKNAHLKQEIAALEQEIAALEQ (SAF-p2D) respectively; or

c) KIAALKQKIAALKQEIDALEYENDALEQ (SAF-p1A) and

d) KI**R**ALKWKNA**H**LKQEIAALEQEIAALEQ (SAF-p2C) respectively; or

e) KIAALKQKIA**S**LKQEIDALEYENDALEQ (SAF-p1C) and

f) KI**R**ALKWKNA**H**LKQEIAALEQEIAALEQ (SAF-p2C) respectively.

According to another aspect of the invention, there is provided a method of producing protein structures, the method comprising providing a mixture of first and second peptide monomer units which associate to form a protein structure according to the invention. The structure can be derivatised and/or stabilized by cross-linking.

The invention also provides protein fibres produced by an association of protein structures according to the invention.

According to another aspect of the invention, there is provided a kit for making protein structures, the kit comprising first and second peptide monomer units which associate to form a protein structure or protein fibres according to the invention.

The protein structures of the invention may be assembled in two and three dimensional arrays. For example, two dimensional mats can be formed which can function, for example

as filters. Three dimensional grids can also be formed again for example, for use as filters or for organising other associated or conjugated molecules in three dimensions.

Functional moieties may be appended to the amino acids of the peptide monomer units. For example, fluorescent moieties (fluorophores) may be attached to the coiled coil as described in WO99/11774. The addition of fluorescent moieties may assist visualization of the protein structure. Substitution with functional groups at the "f" position in the heptad repeat is especially preferred as that position is on the outside of the helix (see Fig. 1C and 1E).

The stability of the protein structures at higher temperatures may be improved by making the peptide monomers longer, such that the overlap between corresponding first and second monomer unit residues is increased. Increases in monomer length have previously been shown to stabilize coiled coil structures. Alternatively, stability can be improved by introducing bonding between adjacent peptide monomer units in the same strand. For example, Kent (Dawson *et al* (1994) Science **266**: 776) and co-workers have produced peptide bonds between adjacent polypeptide units by coupling and subsequent rearrangement of a cysteine residue at the N end of one polypeptide unit to a thio-ester derivatised C-terminus of another unit.

Definitions

The terms used in the specification are to be given the ordinary meaning attributed to them by the skilled addressee. The following is given by way of clarification:

Amino acid.

This term embraces both naturally-occuring amino acids and synthetic amino acids as well as naturally-occuring amino acids which have been modified in some way to alter certain properties such as charge. In all cases references to naturally-occurring amino acids may be considered to include synthetic amino acids which may be substituted therefor.

Coiled Coil

A coiled coil is a peptide/protein sequence with a contiguous pattern of hydrophobic residues spaced 3 and 4 residues apart, which assembles (folds) to a multi-meric bundle of helices.

Dimer

A dimer is a two stranded structure.

Heterodimer

A heterodimer is a dimeric structure formed by two different stands.

Staggered heterodimer

A staggered heterodimer is a structure in which the two strands assemble to leave overlagging ends that are not interacting within the heterodimer.

Protofibril

A protofibril is a protein structure assembled longitudinally from staggered heterodimers interacting through their overhanging ends.

Fibre

A fibre is a structure formed by lateral association of two or more protofibril.

Protein structures and methods of producing protein structures in accordance with the invention will now be described, by way of example only, with reference to the accompanying Figures 1 to 4 in which:

Fig. 1 illustrates the design and the sequences of self-assembling fibre (SAF) peptide monomers of the invention.

Fig. 2 illustrates computer modeling of the designed self-assembling fibre of the invention.

Fig. 3 illustrates the results of circular dichroism (CD) and linear dichroism (LD) experiments on protein structures of the invention.

Fig. 4 illustrates the assembly of synthetic protein fibres visualized directly by transmission electron microscopy and an analysis of fibre width In all panels, the white scale bars represent 100 nm. Fig. 4D is a histogram showing the distribution of fibre widths determined using TEM for fresh (white bars) and matured (black bars) mixtures of SAF peptides at 100 μM (a width value of "*x*" on the histogram includes all measurements made from "(*x*–*5*) to *x*").

1) Peptide Design and Synthesis

Various peptide monomer units were designed as described above. The monomers and capping peptides (designed to complement the sticky ends of the monomers so as to produce flush, or blunt ends and, so, arrest longitudinal fibre assembly) are set out in Table 1:

TABLE 1

PEPTIDE	SEQUENCE g abcdefg abcdefg abcdefg abcdef 5 10 15 20 25	DESIGN	CD DATA		LD DATA		EM DATA	
			@ 10 μM	@ 100 μM	@ 10 μM	@ 100 μM	@ 10 μM	@ 100 μM
CAP-p1A	*[1]YGPGE IAALEQE NAALEQ	prototype	unfolded					
SAF-p1A	K IAALKQK IAALKQE IDALEYE NDALEQ	prototype; slowly precipitates	unfolded	~ 45 % α-helix				no fibres
SAF-p1B	*K IAALKQK IAALKQE IDALEYE NDALEQ*	chemical capping of the ends (↑ stability)	~ 60 % α-helix	~ 70 % α-helix				
SAF-p1C	K IAALKQK IASLKQE IDALEYE NDALEQ	no capping (↓ stability); mutate A_{11} → S (↑ solubility & ↓ helix stability)	unfolded	~ 20 % α-helix		no signal		no fibres
CAP-p2A	K IAALKQK NAALKQG GW*	prototype	unfolded					
SAF-p2A	K ISALKWK NASLKQE IAALEQE IAALEQ	prototype; low solubility	unfolded					
SAF-p2B	*K IRALKWK NAHLKQE IAALEQE IAALEQ*	mutate S_3 → R & S_{11} → H (↑ solubility & ↑ helix stability)	~ 60 % α-helix	~ 95 % α-helix				
SAF-p2C	K IRALKWK NAHLKQE IAALEQE IAALEQ	no capping (↓ stability)	unfolded	~ 20 % α-helix				thin fibres
SAF-p2D	K IRALKAK NAHLKQE IAALEQE IAALEQ	mutate W_7 → A (investigate role of Trp in fibrillogenesis)	~ 15 % α-helix	~ 45 % α-helix		no signal		no fibres

[1] *=Chemical capping = CH_3CO at the N terminus and NH_2 at the C terminus

PEPTIDE	SEQUENCE g abcdefg abcdefg abcdefg abcdef 5 10 15 20 25	DESIGN	CD DATA @ 10 μM	@ 100 μM	LD DATA @ 10 μM	@ 100 μM	EM DATA @ 10 μM	@ 100 μM
SAF-p2E	K IRALKCK NAHLKQE IAALEQE IAALEQ	mutate $A_7 \rightarrow C$ (for derivatization & cross-linking)						

Fig. 1 shows (A) A mechanism for self-assembly: complementary charges in "companion" peptides direct the formation of staggered, parallel heterodimers; the resulting "sticky" ends are also complementary and promote longitudinal association into extended structures. Fig 1(B) shows the designed amino acid sequences: each peptide comprised canonical heptad repeats (***abcdefg***) with Ile at ***a*** and Leu at ***d*** to guide the formation of coiled-coil dimers; oppositely-charged residues were incorporated at ***e*** and ***g*** to favour the staggered dimer with sticky ends; asparagine residues (which preferentially pairs with each other at ***a*** sites (Gonzalez L *et al* (1996) Nature Structural Biology 3, **13**: 1011-1018) were included to cement the prescribed register further and to favoor the parallel structures. Fig. 1(C) is a helical-wheel representation, summarizing the designed sequences in context. The view is from the N-terminus with heptad sites labeled ***a-g*** and assumes 3.5 residues per helical turn to emphasise the heptad repeat.

The peptides were synthesized on an Applied Biosystems 432A Peptide Synthesizer using solid-phase methods and Fmoc chemistry. Peptide samples were purified using reversed-phase HPLC and their identities confirmed by MALDI-TOF mass spectrometry.

Various combinations of peptide monomers and capping peptides were tested as set out in Table 2:

TABLE 2

MISC.	PEPTIDE MIXTURE	EQUILIBRATION	CD DATA @ 10 μM	@ 100 μM	LD DATA @ 10 μM	@ 100 m μM	EM DATA @ 10 μM	@ 100 μM
	SAF-p1A, SAF-p2A		~ 20 % α-helix	~ 40 % α-helix $T_m \approx 30$ °C				
	SAF-p1A, SAF-p2A, CAP-p1A, CAP-p2A			~ 20 % α-helix				
	SAF-p1A, SAF-p2B	rapid		~ 65 % α-helix $T_m \approx 36/46$ °C (expect 38 °C if no interaction)				
	SAF-p1B, SAF-p2B	rapid	~ 70 % α-helix $T_m \approx 25$ °C (expect 19 °C if no interaction)	~ 80 % α-helix $T_m \approx 35$ °C (expect 36 °C if no interaction)				no fibres
	SAF-p1A, SAF-p2C	slow @ 100 μM; clouding occurs	~ 20 % α-helix	~ 80 % α-helix biphasic thermal melt			no fibres	thick fibres (~ 45 nm wide)
no signal from fibres in 1D-NMR[2]	SAF-p1C, SAF-p2C	slow @ 100 μM; clouding occurs	~ 25 % α-helix	~ 65 % α-helix (~25% in 0.5 M salt); unusual spectral shape; no aromatic signal	no signal	strong signal from backbone & aromatics	no fibres	thick fibres (~ 45 nm wide); no fibres in 0.5 M salt
	SAF-p1C, SAF-p2D	slow @ 100 μM; clouding occurs	~ 25 % α-helix	~ 85 % α-helix (~20% in 0.5 M salt); unusual spectral shape	no signal	strong signal from backbone & aromatics; no signal in 0.5 M salt	no fibres	thick fibres (~ 45 nm wide); no fibres in 0.5 M salt

[2] Indicates large structures are formed.

2) Modeling of Protein Fibre Structure

A model of the three-dimensional structure of the designed protein fibre resulting from the assembly of SAF-p1 and SAF-p2 was made from the minimized structure of a model coiled-coil 35-mer, $(LAALAAA)_5$, which was generated using Crick's Equation and had an ideally packed interface (G. Offer and R. Sessions, *J. Mol. Biol.* **249**, 967 (1995)). Copies of the 35-mer were superimposed with an overlap of one heptad repeat to extend the structural template, and the backbone was rejoined after removal of overlapping segments. Residues in the two-stranded template were replaced with the sequences of the SAF peptides, staggered relative to each other by two heptad repeats according to the alignment in Fig. 1B. The structure was soaked in a 5 Å layer of water and energy minimized until the average absolute derivative of coordinates with respect to energy fell below 0.01 kcal $Å^{-1}$. The structure was built and visualized using Insight II 97.0 (Molecular Simulations Inc.), and was energy-minimized using Discover 2.9.8 (Molecular Simulations Inc.) with the consistent valence forcefield. In Fig 2(A) peptides SAF-p1 and SAF-p2 (each coloured dark grey-to-light grey from the N-terminus) interact through core residues including asparagine pairs (coloured mid-grey) to form the two strands of a staggered, parallel, coiled-coil fibre. In Fig. 2(B), negatively charged glutamate side chains (coloured light grey) and positively charged lysine side chains (coloured black) form complementary charge interactions between the SAF peptides.

3) Circular Dichroism Experiments

Peptide samples were incubated at 5°C in 10 mM MOPS (3-(N-Morpholino)propanesulfonic acid), pH 7. Sample concentrations were determined from their UV absorbance at 280 nm (SAF-p1) and 214 nm (SAF-p2). After baseline correction, ellipticities in mdeg were converted to molar ellipticities (deg cm^2 dmol-res^{-1}) by normalizing for the concentration of peptide bonds. Data were recorded in a cell of 1 mm path length by integrating the signal for 5s (and 1s for the fresh 100 μM peptide mixture) every nm in the range 205-260 nm. CD measurements were made using a JASCO J-715 spectropolarimeter fitted with a Peltier temperature controller.

The CD data shown in Fig. 3 provides spectroscopic evidence for the formation of helical structures by the SAF peptides. Fig. 3(A) shows circular dichroism (CD) spectra at 10 μM

for: SAF-p1 (----), SAF-p2 (- -), the average of these spectra (- - -), and the experimental SAF peptide mixture (o). Fig 3(B) shows CD spectra at 100 μM - the key is the same as for Fig 3(A), but with the additional spectrum (•) being for the SAF peptide mixture after "maturation" for 1 h.

Consistent with our design, neither SAF-p1 nor SAF-p2 was highly structured in aqueous solution at pH 7 and 5 °C (Fig. 3). However, when mixed in equal proportions the circular dichroism (CD) spectrum changed and, moreover, was markedly different from the theoretical spectrum generated by averaging the spectra for the isolated peptides. In particular, the spectrum for the mixture had intense minima at 208 and 222 nm consistent with the formation of α-helical structure, but these features were not as pronounced in the spectra of the individual peptides. This was clear evidence that the two peptides interacted to form an α-helical structure as designed. Furthermore, and as expected for a multimerization event, the magnitude of these spectral changes depended on peptide concentration; a SAF mixture with 10 μM of each peptide, did show a weak signal indicative of some α-helical structure, however, a 100 μM mixture gave a much stronger signal (Figs. 3A&B).

The shape and intensity of spectra from 100 μM mixtures of the SAF peptides also changed with time (Fig. 3B). Spectra recorded immediately after mixing a "fresh" sample displayed some α-helical structure. After incubation of the mixture for 1 hour at 5 °C ("maturation"), however, the signal at 222 nm was more intense, and indicated approximately 75 % α-helix, consistent with substantial coiled-coil formation.

Maturation of 100 μM SAF peptide mixtures was also accompanied by slight clouding of the samples. Scattering effects from such samples can lead to attenuation and distortion of CD spectra (D. Mao and B. A. Wallace, (1984) *Biochemistry* **23**, 2667). However, we could disregard this possibility because altering the distance between the sample and the detector in the CD instrument did not affect the shape or the intensity of the spectrum. Furthermore, we established that the majority of the CD signal from the mixtures derived from the suspended material: a supernatant without the suspended material, which was recovered by centrifugation of a matured 100 μM SAF mixture, gave only a weak CD signal similar to the 10 μM mixture.

Thus, the CD data were wholly consistent with the desired α-helical SAF design and, moreover, indicated the formation of large assemblies.

4) Linear Dichroism Experiments

Linear dichroism (LD) spectroscopy was used to test if elongated structures were being formed as designed. Long polymers such as DNA molecules can be oriented by shear flow. This effect can be monitored by LD spectroscopy provided that chromophores also become aligned by the flow (M. Bloemendal (1994) *Chem. Soc. Rev.* **23**, 265; A. Rodger and B. Norden (1997) Oxford Chemistry Masters (Oxford University Press, Oxford), vol. 1).

Peptide samples were prepared for LD as for CD. LD data were collected on samples spinning in a couette flow cell by integrating the signal for 2 s every nm in the range 210-320 nm, using a JASCO J-715 spectropolarimeter. After baseline correction, absorbance was converted to molar extinction coefficient (1 mol-res^{-1} cm^{-1}) by normalizing for the concentration of peptide bonds. A linear correction for a sloping baseline was made to the data from the 100 μM SAF peptide mixture.

The results are depicted in Fig. 3D, which shows linear dichroism (LD) spectra for: 20 μM tropomyosin (- - -), the SAF peptide mixture at 10 μM (----), and the SAF peptide mixture at 100 μM in the absence (•) and presence (o) of 0.5 M KF.

For instance, we found that tropomyosin, which forms a dimeric coiled coil approximately 42 nm in length, could be aligned to give a LD signal (Fig. 3D). In contrast and consistent with our design and the CD data, experiments with a 10 μM SAF mixture, (Fig. 3D), and for the individual peptides at 100 μM (data not shown), LD signals were not detected. However, a matured 100 μM SAF peptide mixture gave a strong absorbance from the peptide backbone (210-240 nm) and some signal in the aromatic region (260-290 nm) during flow orientation (Fig. 3D). As only long structures are aligned by this technique, the data demonstrated that long fibres at least 500 nm in length were present in solutions of the matured 100 μM SAF peptide mixtures.

5) Electron Microscopy

To confirm fibre assembly, we used electron microscopy to visualize structures in the peptide preparations directly. For TEM experiments, peptide samples were incubated for 1 h at 5 °C in filtered 10 mM MOPS, pH 7. A drop of peptide solution was applied to a carbon-coated copper specimen grid (Agar Scientific Ltd, Stansted, UK), and dried with filter paper before negative staining with 0.5% aqueous uranyl acetate and then dried at 5 °C. A "fresh" SAF peptide mixture was prepared by mixing preincubated solutions of the individual peptides at 200 μM directly on the specimen grid, before drying and negative staining as described. Grids were examined in a Hitachi 7100 TEM at 100 kV and digital images were acquired with a (800 x 1200 pixel) charge-coupled device camera (Digital Pixel Co. Ltd., Brighton, UK) and analyzed (Kinetic Imaging Ltd., Liverpool, UK).

For scanning electron microscopy (SEM) experiments, negatively-stained specimen grids were sputter-coated with gold and examined in a Leo Stereoscan 420 SEM at 20 kV and with a probe current of 10 pA.

No structures were visible up to 100 000 times magnification by transmission electron microscopy (TEM) for either the 10 μM SAF mixture, or for the individual peptides at 100 μM concentration (data not shown). However, TEM of a 100 μM SAF mixture at 50 000 times magnification revealed time-dependent formation of long fibrous structures, consistent with the CD and LD data. Fresh mixtures showed large numbers of extended fibres of various widths. The majority of these had a diameter of about 20 nm (Figs. 4A (a fresh mixture at 100 μM) & Fig 4D); finer fibres were present, but their widths could not be measured reliably. Images recorded for the matured mixtures showed fewer fibres, but these were more distinct and thicker than those observed in the fresh mixture (Fig. 4B&D). Scanning electron microscopy (SEM) of a matured mixture showed no evidence for fibre branching. Rather, the fibres were simply intertwined as if layered on top of each other (Fig. 4C). It was not possible to follow the full length of fibres due to intertwining, but they were at least several hundred microns in length. Although the density of fibres varied across the surface of the EM grid, for the matured samples at least, their diameters were quite uniform with a mean width of 43.3 (SD = 9.3) nm (Fig. 4D). As the original design was for a longitudinally extended, but otherwise two-stranded coiled coil the average diameter that we might have expected was about 3 nm. Therefore, the EM data suggested that the designed two-stranded coiled-coil fibres associates laterally into higher order assemblies.

6) Effect of Potassium Fluoride on Protein Fibre Assembly

Molecular modeling of the SAF sequences into an extended two-stranded coiled coil also highlighted potential complementary charge interactions on the surface of the protofibrils, Figs 1&2. In accord with this, experimentally it was found that moderate concentrations of salt inhibited protofibril and thick fibre assembly. First, CD spectra recorded for both the individual peptides and a 100 μM mixture of SAF peptide samples with 0.5 M potassium fluoride showed reduced helical CD signals and there was no evidence of "maturing" in the mixed samples (Fig. 3C). Second, the LD signal described previously for the matured 100 μM SAF peptide mixture was also lost when the experiment was repeated in the presence of salt (Fig. 3D). Finally, TEM images of a 100 μM SAF mixture also demonstrated that fibres were not formed in the presence of 0.5 M KF (Fig. 4E). Fig. 4E shows the results of TEM of a matured SAF peptide mixture at 100 μM incubated in the presence of 0.5 M KF.

The inventors did not knowingly design any features into the SAF peptides to foster further association of the two-stranded coiled coils. The observation of thick fibres in SAF peptide preparations, therefore, raised the question: what interactions guided and stabilized these higher-order assemblies? The inventors therefore propose that features inherent in repeating structures of the type that they designed will naturally promote such fibre assembly (fibrillogenesis).

Consider a protofibril as depicted in Fig. 1B and 2A. Any sequence feature presented on its surface by either, or both of the constituent peptides will be repeated at regular intervals along the protofibril. The repeat length will be equal to the length of the peptides (for SAF-p1 and SAF-p2 this was 28 residues, or about 4.2 nm). Furthermore, the motif will spiral around the protofibril tracking the superhelix of the coiled coil, which has a pitch of about 15 nm for a contiguous, heptad-based, dimeric structure. In this scenario, protofibril-protofibril interactions may be promoted if another sequence motif complementary to the first is present in the potential partner. This is because the pitches of the complementary motifs on each protofibril will match precisely. Thus, once initiated, lateral association of protofibrils — that is, fibrillogenesis — will be cemented by many regularly spaced interactions as in a crystal. As a result, the complementary interactions need only be weak as the stability of the protofibril-protofibril interaction rests on an avidity effect rather than a small number of strong interactions. Provided that the components of the assembly can make more than one type of complementary surface very extensive molecular assemblies may result.

The inventors used electrostatic interactions both to direct heterodimer formation, and to promote elongation of the protofibrils (Figs. 1 and 2). These features would also create periodic and alternating patches of charge in the protofibrils provided they are regular as envisaged (Fig. 1B and 2B). These charged patches could guide and stabilize the higher order assemblies. Indeed, similar features have been noted in several natural fibrous proteins and have been implicated in the assembly of multi-protein filaments (J. J. Meng *et al* (1994) *Biol. Chem.* **269**, 18679; A. D. McLachlan and M. Stewart (1976) *Mol. Biol.* **103**, 271), and small synthetic peptide systems (S. G. Zhang *et al* (1993) *Proc. Natl. Sc. U.S.A* **90**, 3334). The experiments with salt (KF) described above suggest that salt-bridges (electrostatic interaction) may be at least in part the cause of fibrillogenesis.

Claims

1. A protein structure comprising a plurality of first peptide monomer units arranged in a first strand and a plurality of second peptide monomer units arranged in a second strand in which a first peptide monomer unit in the first strand extends beyond a corresponding second peptide monomer unit in the second strand in the direction of the strands.
2. A protein structure according to claim 1 in which the strands together form a coiled coil structure.
3. A protein structure according to claim 1 or 2 in which at least one charged amino acid residue of a first peptide monomer unit is arranged to attract an oppositely-charged amino acid residue of a second peptide monomer unit.
4. A protein structure according to claim 3 in which the charged amino acid residue is in an end portion of the first peptide monomer unit which extends beyond the corresponding second peptide monomer unit in the second strand.
5. A polypeptide structure according to any preceding claim in which at least one strand consists solely of first or second peptide monomer units respectively.
6. A protein structure according to any preceding claim in which the peptide monomer units comprise a repeating structural unit.
7. A protein structure according to claim 6 in which the repeating structural unit comprises a heptad repeat motif (abcdefg).
8. A protein structure according to claim 6 in which the repeating structural unit comprises a hendecad repeat motif (abcdefghijk)
9. A protein structure according to claim 6 having isoleucine or asparagine at position a and leucine at position d.
10. A protein structure according to claim 6 having valine or leucine at positions a and d respectively.
11. A protein structure according to any one of claims 7 to 10 having oppositely-charged or otherwise complementary residues at positions g and e of respective monomer units.
12. A protein structure according to claim 11 in which the oppositely-charged residues are glutamic acid and lysine residues or asparagine and aspartic acid, or synthetic derivatives.

13. A protein structure according to any preceding claim in which the structure is stabilised by pairs of asparagine, arginine, lysine or other complementary residues provided by corresponding first and second peptide monomer units.

14. A protein structure according to any preceding claim in which the first and second peptide monomer units have the sequence:

a) KIAALKQKIASLKQEIDALEYENDALEQ (SAF-p1) and

b) KIRALKAKNAHLKQEIAALEQEIAALEQ (SAF-p2) respectively; or

c) KIAALKQKIAALKQEIDALEYENDALEQ (SAF-p1A) and

d) KI**R**ALKWKNA**H**LKQEIAALEQEIAALEQ (SAF-p2C) respectively; or

e) KIAALKQKIA**S**LKQEIDALEYENDALEQ (SAF-p1C) and

f) KI**R**ALKWKNA**H**LKQEIAALEQEIAALEQ (SAF-p2C) respectively.

15. A peptide monomer unit for use in preparing a protein structure the peptide monomer unit having an amino acid sequence selected from:

a) KIAALKQKIASLKQEIDALEYENDALEQ (SAF-p1);

b) KIRALKAKNAHLKQEIAALEQEIAALEQ (SAF-p2);

c) KIAALKQKIAALKQEIDALEYENDALEQ (SAF-p1A);

d) KI**R**ALKWKNA**H**LKQEIAALEQEIAALEQ (SAF-p2C);

e) KIAALKQKIA**S**LKQEIDALEYENDALEQ (SAF-p1C) ; and

d) KI**R**ALKWKNA**H**LKQEIAALEQEIAALEQ (SAF-p2C).

16. A method of producing protein structures, the method comprising providing a mixture of first and second peptide monomer units which associate to form a protein structure according to any preceding claim.

17. A method according to claim 16 in which the protein structure is derivatised.

18. A method according to claim 16 or 17 in which the protein structure is stabilised by cross-linking.

19. Protein fibres produced by an association of protein structures according to any one of claims 1 to 3 or a method according to claim 16, 17 or 18.

20. A kit for making protein structures, the kit comprising first and second peptide monomer units which associate to form a protein structure according to any one of claims 1 to 13 or protein fibres according to claim 19.

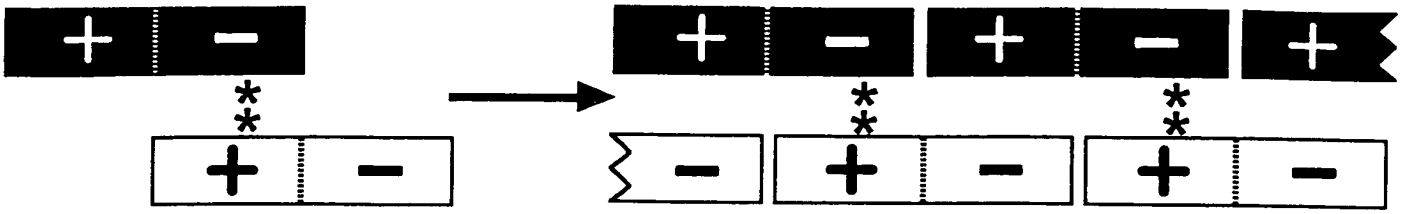
21. A two dimensional matt comprising a protein structure according to any one of claims 1 to 13 or protein fibres according to claim 19.

22. A three dimensional grid comprising a protein structure according to any one of claims 1 to 13 or protein fibres according to claim 19.

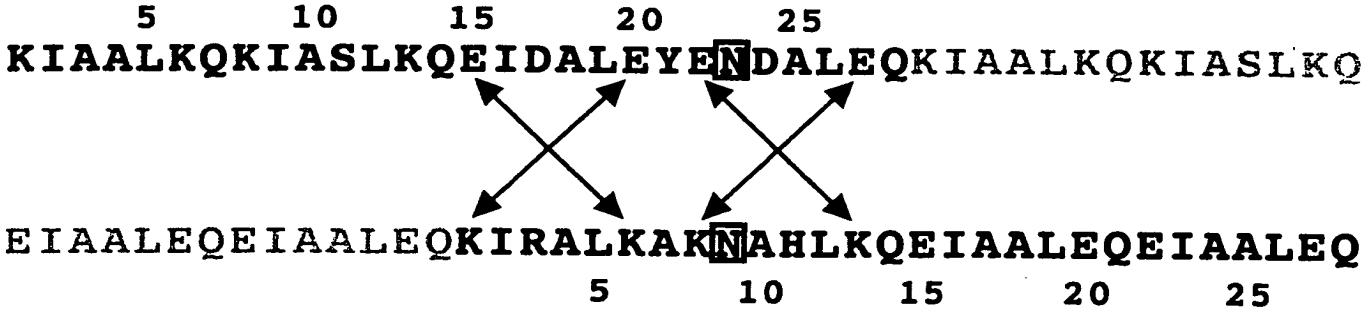
Protein fibre formation

This invention relates to protein fibre formation and in particular to methods of producing protein fibres to form a protein structure comprising a plurality of first polypeptide units arranged in a first polypeptide strand and a plurality of second polypeptide units arranged in a second polypeptide strand, the strands preferably forming a coiled coil structure, and in which a first polypeptide unit in the first strand extends beyond a corresponding second polypeptide unit in the second strand in the direction of the strands.

FIG. 1



B


5 10 15 20 25
KIAALKQKIASLKQEIDALEYENDALEQKIAALKQKIASLKQ
EIAALEQEIAALEQKIRALKAKNAHLKQEIAALEQEIAALEQ
5 10 15 20 25

C

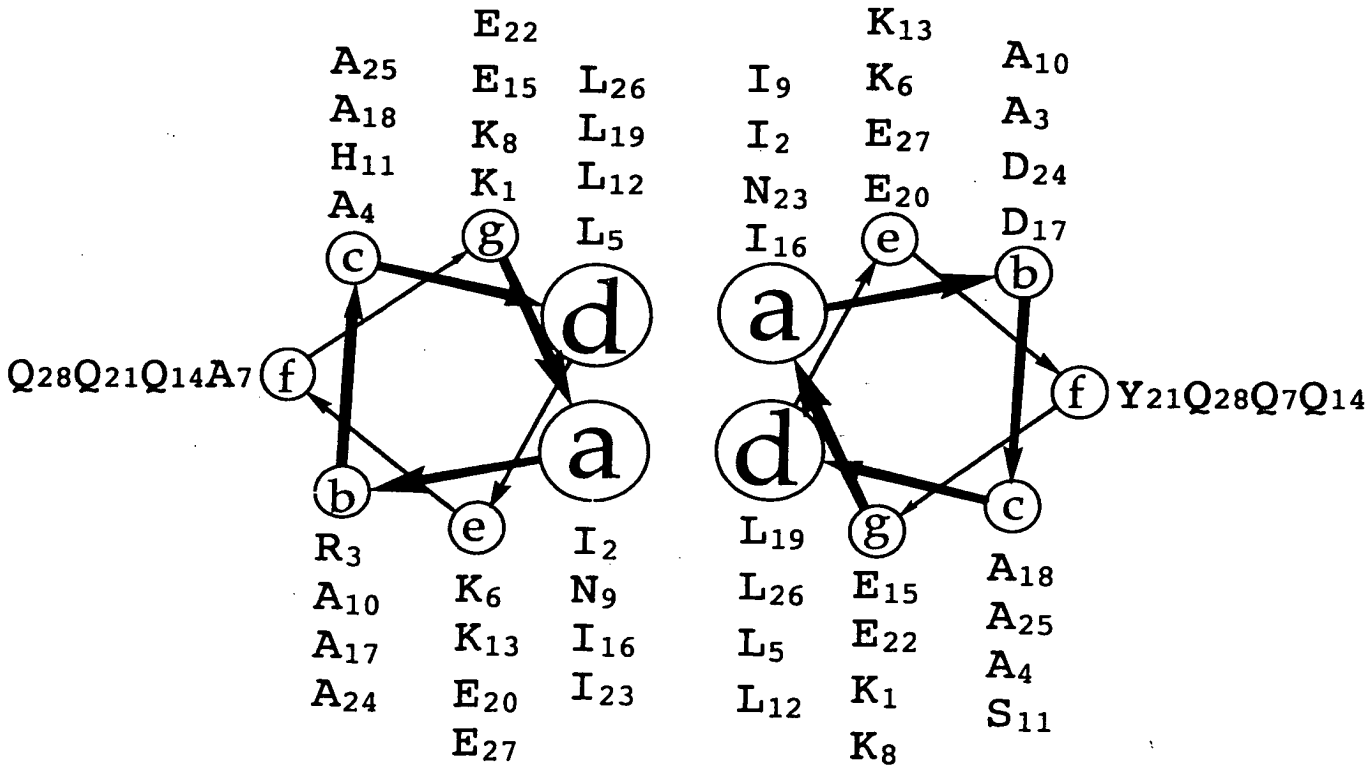

E22
A25 E15 L26
A18 K8 L19
H11 K1 L12
A4 L5
Q28Q21Q14A7
R3 I2
A10 K6 N9
A17 K13 I16
A24 E20 I23
E27
K13
I9 K6 A10
I2 E27 A3
N23 E20 D24
I16 D17
Y21Q28Q7Q14
L19 A18
L26 E15 A25
L5 E22 A4
L12 K1 S11
K8

FIG. 2

Fig 2A

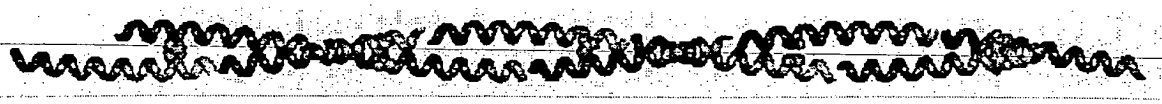


Fig. 2B

FIG. 3

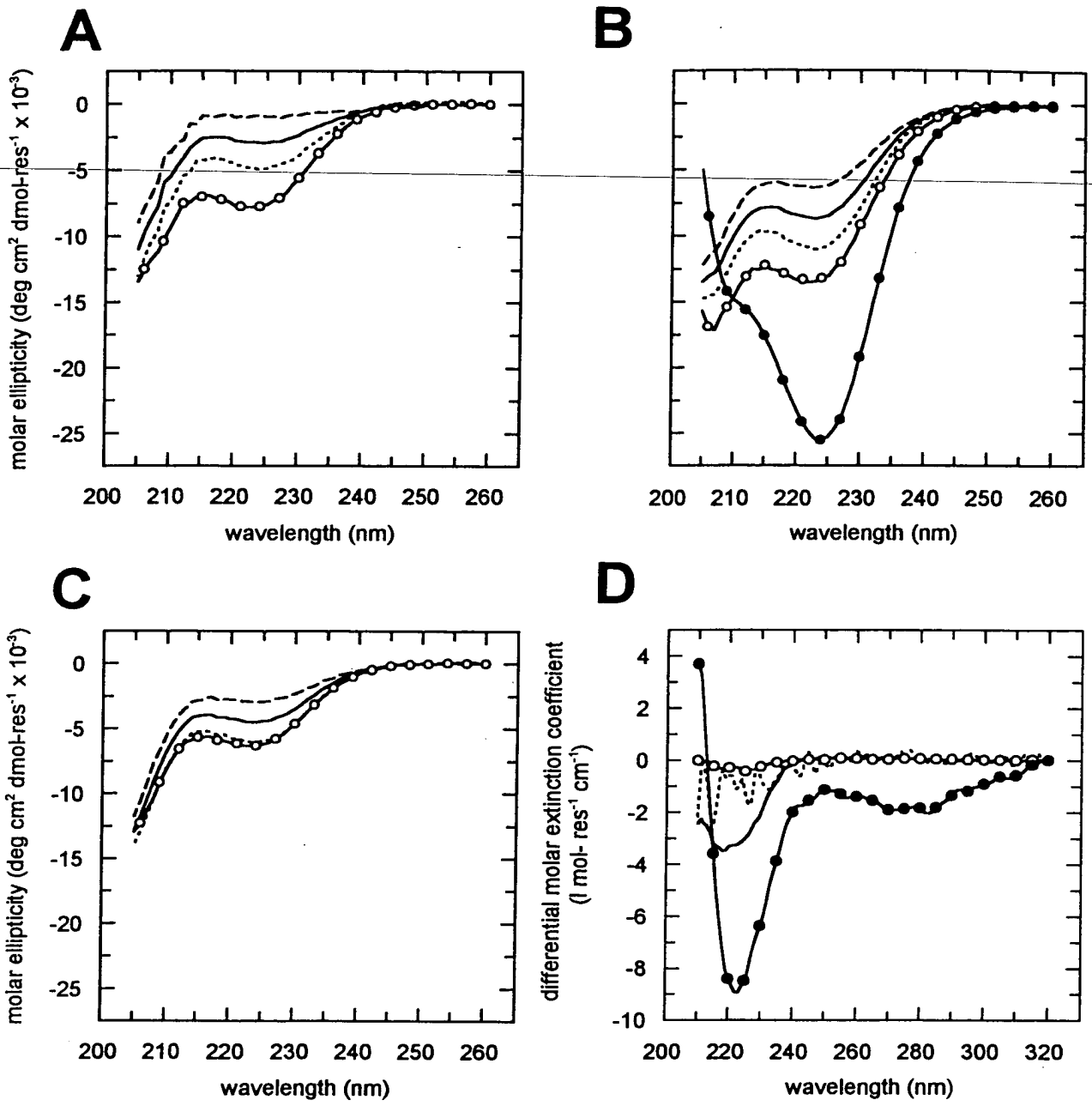
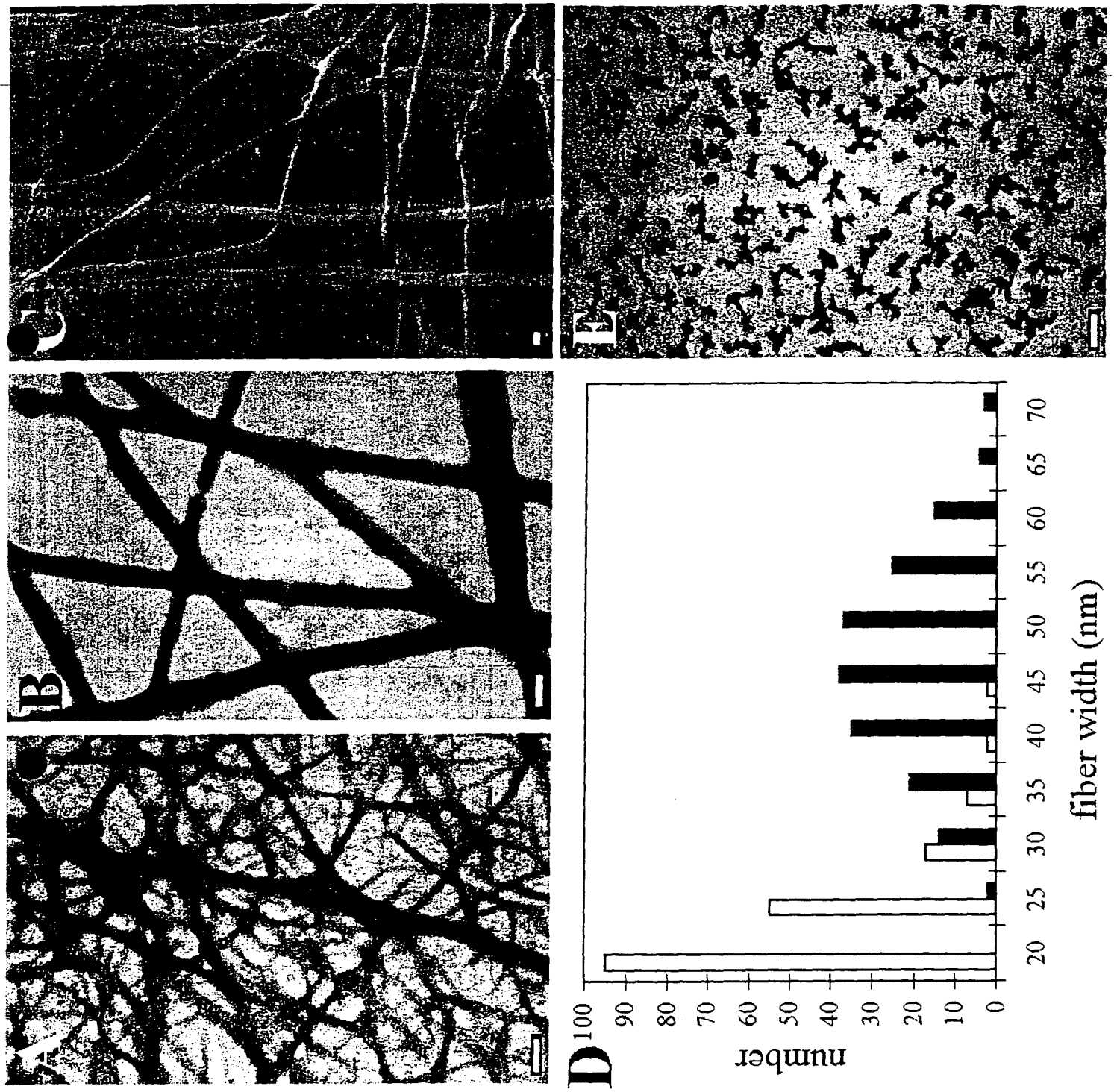

A
B
C
D
molar ellipticity (deg cm² dmol-res⁻¹ x 10⁻³)
differential molar extinction coefficient (l mol- res⁻¹ cm⁻¹)
wavelength (nm)
0
-5
-10
-15
-20
-25
4
2
-2
-4
-6
-8
200
210
220
230
240
250
260
280
300
320

FIG. 4


A
B
C
D
E
number
fiber width (nm)
0
10
20
30
40
50
60
70
80
90
100
20
25
30
35
40
45
50
55
60
65
70

PCT NO: GB00/03576

Fraser Davies